



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Reprint: Good laboratory practice: preventing introduction of bias at the bench

**Citation for published version:**

Macleod, MR, Fisher, M, O'Collins, V, Sena, ES, Dirnagl, U, Bath, PMW, Buchan, A, van der Worp, HB, Traystman, RJ, Minematsu, K, Donnan, GA & Howells, DW 2009, 'Reprint: Good laboratory practice: preventing introduction of bias at the bench', *Journal of Cerebral Blood Flow and Metabolism*, vol. 29, no. 2, pp. 221-223. <https://doi.org/10.1038/jcbfm.2008.101>

**Digital Object Identifier (DOI):**

[10.1038/jcbfm.2008.101](https://doi.org/10.1038/jcbfm.2008.101)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Journal of Cerebral Blood Flow and Metabolism

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



Published in final edited form as:

*J Cereb Blood Flow Metab.* 2009 February ; 29(2): 221–223. doi:10.1161/STROKEAHA.108.525386.

## Reprint: Good laboratory practice: preventing introduction of bias at the bench

Malcolm R Macleod<sup>1</sup>, Marc Fisher<sup>2</sup>, Victoria O'Collins<sup>3,4</sup>, Emily S Sena<sup>1,3,4</sup>, Ulrich Dirnagl<sup>5</sup>, Philip MW Bath<sup>6</sup>, Alistair Buchan<sup>7</sup>, H Bart van der Worp<sup>8</sup>, Richard J Traystman<sup>9</sup>, Kazuo Minematsu<sup>10</sup>, Geoffrey A Donnan<sup>3,4</sup>, and David W Howells<sup>3,4</sup>

<sup>1</sup>Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK <sup>2</sup>Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts, USA <sup>3</sup>National Stroke Research Institute, Austin Health, University of Melbourne, Victoria, Australia <sup>4</sup>Department of Medicine, Austin Health, University of Melbourne, Melbourne, Victoria, Australia <sup>5</sup>Charité Department for Experimental Neurology, Center for Stroke Research Berlin, Berlin, Germany <sup>6</sup>Stroke Trials Unit, University of Nottingham, Nottingham, UK <sup>7</sup>Acute Stroke Program, Nuffield Department of Medicine, John Radcliffe Hospital, Oxford, UK <sup>8</sup>Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center, Utrecht, The Netherlands <sup>9</sup>Anschutz Medical Campus, University of Colorado Denver, Aurora, Colorado, USA <sup>10</sup>Cerebrovascular Division, Department of Medicine, National Cardiovascular Center, Osaka, Japan

### Abstract

As a research community, we have failed to show that drugs, which show substantial efficacy in animal models of cerebral ischemia, can also improve outcome in human stroke. Accumulating evidence suggests this may be due, at least in part, to problems in the design, conduct, and reporting of animal experiments which create a systematic bias resulting in the overstatement of neuroprotective efficacy. Here, we set out a series of measures to reduce bias in the design, conduct and reporting of animal experiments modeling human stroke.

### Keywords

animal models; drug trials; experimental; outcomes; preventing bias; translational research

Nearly 10 years after the first Stroke Therapy Academic Industry Roundtable (STAIR) participants established guidelines intended to support the translation of neuroprotective efficacy from bench to bedside (Altman *et al*, 2001), there is still no clinically effective neuroprotective drug for stroke. One interpretation of this observation is that measures outlined in STAIR I have failed to deliver the promised improvements in drug development. However, a dispassionate analysis of data presented over the last 10 years suggests that the 'STAIR hypothesis'—that improvements in animal experimental design will lead to improvements in translational efficiency—has yet to be adequately tested. Adhering to standards of conducting and reporting of experiments to reduce the confounding effects of bias and ensure adequate statistical power, as outlined below, will increase the confidence with which we can assess new data and maximize our chances of developing effective therapies.

The original STAIR proposal was that by paying due attention to experimental bias, to the breadth of physiologic variables known to influence stroke outcome in patients, and by testing therapies in a range of model systems which might more faithfully reproduce the key facets of stroke pathophysiology, we would be able to translate what appeared to be clear evidence of neuroprotective efficacy in animals to the more heterogeneous circumstances of human stroke. Although we believe strongly that failure to adequately consider variables such as age, comorbidity, physiologic status, and timing of drug administration contribute to the disparity between the results of animal models and clinical trials, they have been reviewed elsewhere (Altman *et al*, 2001; Bath *et al*, 1998) and are not the subject of this article.

Analyses of data supporting the efficacy of various neuroprotective strategies (Begg *et al*, 1996; Crossley *et al*, 2008; Dirnagl, 2006) have revealed that although many researchers adhere closely to the ethos of these guidelines, as a community we do not. A simple checklist derived from the STAIR guidelines to provide an overview of the range of data available for 1,026 candidate therapies (Crossley *et al*, 2008) revealed that only a few came close to meeting the STAIR guidelines. A higher score against this checklist was accompanied by a marked reduction in effect size. This later trend could be seen clearly even within the data for individual drugs (Grotta, 1995). Moreover, studies which reported measures to avoid bias such as random allocation to treatment group, masked induction of ischemia, or the masked assessment of outcome (Macleod *et al*, 2005, 2008), gave a markedly lower estimate of efficacy. Despite this there has been some evidence of improvement in study quality, and the performance of animal stroke studies is substantially better than that for most other models of neurologic disease (Dirnagl, 2006). And yet, the majority of investigators still do not report whether they took measures to avoid bias.

Systematic reviews and meta-analyses of data from animal stroke studies suggest that these studies may be substantially distorted by experimental bias. Taken together, publications supporting the efficacy of NXY-059 include randomized data with allocation concealment and masked outcome assessment, but most individual publications do not report these measures. Analyses of those data suggest that at least half of the reported 44% improvement in outcome could be attributed to experimental bias, specifically a failure to randomize the allocation to experimental group, a failure to conceal treatment group allocation from the surgeon or a failure to blind the assessment of outcome (Macleod *et al*, in press). Similar observations have been made of the hypothermia literature, where nonrandomized studies and studies without masked outcome assessment appear to give a relative overstatement of efficacy of 27% and 19%, respectively (Macleod *et al*, 2005). Despite the widely recognized importance of these aspects of study design, analyses conducted by the collaborative approach to meta analysis and review of animal data from experimental stroke (CAMARADES) group suggest that only 36% of studies reported random allocation to treatment group, only 11% report allocation concealment, and only 29% reported the masked assessment of outcome (Dirnagl, 2006).

A related issue is the number of animals used in experiments. The probability of detecting a difference of a given size between groups is related to the number of animals in each group, the size of the difference and the variability in the outcome measure used. However, only 3% of studies identified in systematic reviews reported using a sample size calculation (Dirnagl, 2006). Importantly, if sample size calculations are based on falsely large estimates of effect size, studies will not be powered to detect real differences between treatment and control groups. Indeed, *post hoc* analysis suggests that most experimental stroke studies have only a one in three chance of detecting a 20% difference in outcome.

These problems are not unique to the preclinical study of stroke. Clinical stroke trials have had problems with inadequate sample size (O'Collins *et al*, 2006) and have also failed to report whether they took measures to avoid bias (Plint *et al*, 2006). Indeed Cochrane's observation

that ‘when humans have to make observations there is always the possibility of bias’ (Sena *et al*, 2007) was a lynchpin of the CONSORT (*Consolidated Standards of Reporting Trials*) initiative to improve the reporting, design, conduct, analysis, and interpretation of randomized controlled trials to inform decision making in health care (Sena *et al*, 2007; STAIR, 1999). This initiative led to substantial improvements in the reporting and conduct of clinical trials (van der Worp *et al*, 2007).

On the basis of the available evidence it would now seem reasonable to suggest that preclinical testing in animal models of stroke, and indeed other models of disease, should adopt similar standards to ensure that decision making is based on high quality unbiased data (Dirnagl, 2006; Weaver *et al*, 2004). Adoption of such standards would have the added benefit of reducing wasteful usage of financial and animal resources.

In general, studies should only be considered for publication if their ‘Methods’ section includes a description of how they have addressed the standards below, or if authors make a cogent argument for why these standards are not relevant to their work. For these components of a paper, citation of methods described in previous publications is not considered sufficient. These requirements should not preclude publication of important observational, pilot or hypothesis-generating data, but the conclusions of such studies should reflect their preliminary nature.

### **(1) Animals**

The precise species, strain, substrain, and source of animals used should be stated. Where applicable (for instance in studies with genetically modified animals) the generation should also be given, as well as the details of the wild-type control group (for instance littermate, back cross, etc.).

### **(2) Sample size calculation**

The manuscript should describe how the size of the experiment was planned. If a sample size calculation was performed this should be reported in detail, including the expected difference between groups, the expected variance, the planned analysis method, the desired statistical power, and the sample size thus calculated. For parametric data, variance should be reported as 95% confidence limits or standard deviations rather than as the standard error of the mean.

### **(3) Inclusion and exclusion criteria**

Where the severity of ischemia has to reach a certain threshold for inclusion (for instance a prespecified decrease in perfusion detected with laser-Doppler flowmetry, or the development of neurologic impairment of a given severity) this should be stated clearly. Usually, these criteria should be applied before the allocation to experimental groups. If a prespecified lesion size is required for inclusion this should be detailed, as well as the corresponding exclusion criteria.

### **(4) Randomization**

The manuscript should describe the method by which animals were allocated to experimental groups. If this allocation was by randomization, the method of randomization (coin toss, computer-generated randomization schedules) should be stated. Picking animals ‘at random’ from a cage is unlikely to provide adequate randomization. For comparisons between groups of genetically modified animals (transgenic, knockout), the method of allocation to for instance sham operation or focal ischemia should be described.

## **(5) Allocation concealment**

The method of allocation concealment should be described. Allocation is concealed if the investigator responsible for the induction, maintenance, and reversal of ischemia and for decisions regarding the care of (including the early killing of) experimental animals, has no knowledge of the experimental group to which an animal belongs. Allocation concealment might be achieved by having the experimental intervention administered by an independent investigator, or by having an independent investigator prepare drug individually and label it for each animal according to the randomization schedule as out-lined above. These considerations also apply to comparisons between groups of genetically modified animals, and if phenotypic differences (e.g., coat coloring) prevent allocation concealment this should be stated.

## **(6) Reporting of animals excluded from analysis**

All randomized animals (both overall and by treatment group) should be accounted for in the data presented. Some animals may, for very good reasons, be excluded from analysis, but the circumstances under which this exclusion will occur should be determined in advance, and any exclusion should occur without knowledge of the experimental group to which the animal belongs. The criteria for exclusion and the number of animals excluded should be reported.

## **(7) Masked assessment of outcome**

The assessment of outcome is masked if the investigator responsible for measuring infarct volume, for scoring neurobehavioral outcome or for determining any other outcome measures has no knowledge of the experimental group to which an animal belongs. The method of blinding the assessment of outcome should be described. Where phenotypic differences prevent the masked assessment of for instance neurobehavioral outcome, this should be stated.

## **(8) Reporting potential conflicts of interest and study funding**

Any relationship which could be perceived to introduce a potential conflict of interest, or the absence of such a relationship, should be disclosed in an acknowledgements section, along with information on study funding and for instance supply of drugs or of equipment.

We consider that these measures are of central importance to Good Laboratory Practice in the modeling of cerebral ischemia. Many groups already perform experiments to these high standards, and we hope that they will now report this in full, and that others follow their lead. Finally, we do not consider these requirements to represent a final or complete list of appropriate measures necessary to avoid bias. Future additions may be required as further evidence emerges and the experience of authors and reviewers evolves.

## **References**

- Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T. The revised consort statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 2001;134:663–694. [PubMed: 11304107]
- Bath FJ, Owen VE, Bath PM. Quality of full and final publications reporting acute stroke trials: A systematic review. *Stroke* 1998;29:2203–2210. [PubMed: 9756604]
- Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials. The consort statement. *JAMA* 1996;276:637–639. [PubMed: 8773637]

- Crossley NA, Sena E, Goehler J, Horn J, van der Worp B, Bath PM, Macleod M, Dirnagl U. Empirical evidence of bias in the design of experimental stroke studies: a metaepidemiologic approach. *Stroke* 2008;39:929–934. [PubMed: 18239164]
- Dirnagl U. Bench to bedside: the quest for quality in experimental stroke research. *J Cereb Blood Flow Metab* 2006;26:1465–1478. [PubMed: 16525413]
- Grotta J. Why do all drugs work in animals but none in stroke patients? 2. Neuroprotective therapy. *J Intern Med* 1995;237:89–94. [PubMed: 7830036]
- Macleod MR, O'Collins T, Horky LL, Howells DW, Donnan GA. Systematic review and metaanalysis of the efficacy of fk506 in experimental stroke. *J Cereb Blood Flow Metab* 2005;25:713–721. [PubMed: 15703698]
- Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA. Evidence for the efficacy of nxy-059 in experimental focal cerebral ischaemia is confounded by study quality. *Stroke*. 2008advance online publication 14 August 2008
- O'Collins VE, Macleod MR, Donnan GA, Horky LL, van der Worp BH, Howells DW. 1,026 experimental treatments in acute stroke. *Ann Neurol* 2006;59:467–477. [PubMed: 16453316]
- Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, Gaboury I. Does the consort checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust* 2006;185:263–267. [PubMed: 16948622]
- Sena E, van der Worp HB, Howells D, Macleod M. How can we improve the pre-clinical development of drugs for stroke? *Trends Neurosci* 2007;30:433–439. [PubMed: 17765332]
- STAIR. Recommendations for standards regarding preclinical neuroprotective and restorative drug development. *Stroke* 1999;30:2752–2758. [PubMed: 10583007]
- van der Worp HB, Sena ES, Donnan GA, Howells DW, Macleod MR. Hypothermia in animal models of acute ischaemic stroke: a systematic review and metaanalysis. *Brain* 2007;130:3063–3074. [PubMed: 17478443]
- Weaver CS, Leonardi-Bee J, Bath-Hextall FJ, Bath PM. Sample size calculations in acute stroke trials: a systematic review of their reporting, characteristics, and relationship with outcome. *Stroke* 2004;35:1216–1224. [PubMed: 15031455]